Personal storytelling

Tintarev, Nava; Reiter, Ehud; Black, Rolf; Waller, Annalu; Reddington, Joe

Published in:
International Journal of Human-Computer Studies

DOI:
10.1016/j.ijhcs.2016.04.005

Publication date:
2016

Personal Storytelling: Using Natural Language Generation for Children with Complex Communication Needs, in the Wild...

Nava Tintarev[a,*], Ehud Reiter[a], Rolf Black[b], Annalu Waller[b], Joe Reddington[c]

[a]*Department of Computing Science, University of Aberdeen, Aberdeen, AB24 3UE, UK*
[b]*School of Science and Engineering, University of Dundee, Dundee, DD1 4HN, UK*
[c]*eQuality Time Ltd, 12 Canterbury Mansions, London, NW6 1SE, UK*

Abstract

This paper describes a Natural Language Generation system (NLG), How was School Today? that automatically creates a personal narrative from sensor data and other media (photos and audio). It can be used by children with complex communication needs in schools to support interactive narrative about personal experiences. The robustness of story generation to missing data was identified as a key area for improvement in a feasibility study of the system at a fi special needs school. This paper therefore suggests three possible methods for generating stories from unstructured data: clustering by voice recording, by location, or by time. Clustering based on voice recordings resulted in stories that were perceived as most easy to read, and to make most sense, by parents in a quantitative evaluation. This method was implemented in the live system, which was developed and evaluated iteratively at a second special needs school with children with diff t usage profi Open challenges and possibilities for nlg in augmented and alternative communication are also discussed.

Keywords: Assistive technology, Augmented and Alternative Communication, Natural Language Generation, Communication Aids, User-centered Design

---

*Corresponding author
Email address: ntintarev@bournemouth.ac.uk (Nava Tintarev)

1. Introduction

Natural Language Generation (nlg) systems automatically create understandable texts in English or other human languages. These typically start from non-linguistic representations of information as input, and use knowledge about language and the application domain to automatically produce these texts [1]. In this paper, we describe how nlg technology has been used to help users create stories from lifelog data; in particular for children with complex communication needs (CCN). The term CCN describes individuals who, due to motor, language, cognitive, and/or sensory perceptual impairments (e.g., as a result of cerebral palsy), do not develop speech and language skills as expected. This heterogeneous group typically experiences restricted access to the environment, limited interactions with their communication partners, and few opportunities for communication [2].

Electronic Augmentative and Alternative Communication (AAC) systems enable individuals with CCN to verbally communicate their needs, often using Text-to-Speech technology, using single words or simple sentences. AAC is a novel domain for Natural Language Generation, and this work is diff t from much of the previous nlg work on narrative, which has focused on automatically summarising technical data to help expert users perform clearly defi tasks (such as decision making or handover) [3, 4, 5]. In contrast to the above, the nlg in this paper focuses on generating phrases to support individuals with little or no functional speech to engage in interactive conversation about everyday events, primarily to enhance social interaction. Also, the system described in this paper diff from a stand-alone report gener- ator as it is part of an authoring tool. Since the goal is supporting social interaction, the nlg system supports a child in generating, and interactively telling, a story about his or her day. This work also diff from research on fictional story generation in the computational creativity community [6]. Since (logged) personal stories are constrained to communicate real-world data, the system cannot develop original content in order to enhance the story.

This paper makes two key contributions, aimed to resolve the issue of generating data from unstructured and possibly limited data. The fi is a quantitative evaluation of algorithms for constructing a coherent personal narrative from raw input data (together with a small amount of contextual knowledge). The second contribution, is a qualitative evaluation of the wider How was School Today? project.

In the next section, we look at related work in assistive technology, the importance of personal narrative, and narrative in Natural Language Generation. In Section 3, we describe the narrative generation system: How was School Today?. We start with a short overview of a feasibility study with the fi prototype in a fi special needs school. This overview is kept brief, as it is reported elsewhere [7], but is used to highlight the changes incorporated into the current version of the system. The changes among other things address a key technical challenge of how to generate stories from unstructured and occasionally limited sensor data. This challenge is addressed in Section 4, where we present three diff t methods for segmenting data into distinct events. Section 5 presents an evaluation of the diff t methods: 26 parents evaluated the resulting stories on their degree of realism, completeness and readability. The best performing method (clustering around voice recordings) was implemented in the system that was developed incrementally at a second special needs school. Section 6 describes a qualitative evaluation of the improved system (using clustering around voice recordings) over 6 months with two children, staff, and parents at the second special needs school. We conclude with an identification of open challenges in Section 7.

2. Related work

In this section, we discuss related work in the domains of Alternative and Augmentative Communication, the importance of personal narrative, and narrative in Natural Language Generation.

2.1. Alternative and Augmentative Communication (AAC)

Although AAC includes equipment such as simple switches, communication boards and other, 'low-tech' approaches, our project concentrates on the use of 'high-tech' communication devices. High-tech AAC devices generally consist of a computer or tablet[1] running software that allows users to type in text to be spoken out loud, or provides access to pre-stored words and phrases, using multi-meaning symbol sequences (Minspeak) or a hierarchical structure of category pages (dynamic screens). The vocabulary content can be edited, but in reality this level of technical skill cannot be expected from most carers and family.

[1] This may be as simple as an iPad or a specially designed system with features like waterproofing, support for additional access methods, and impact protection.

Advances in text-to-speech technology and mobile computing have made a large range of Augmented and Alternative Communication (AAC) devices available to the public. The launch of the iPad in 2010 severely disrupted the marketplace [8]. Internet connected devices such as iPads and other tablets now have significant market penetration. For example, British National Health Service (NHS) Trusts buy more iPads for AAC than any other equivalent device [9].

Current commercial AAC tools are however poor at supporting storytelling. Basic storytelling can be achieved using 'low-tech' AAC by recording a sequence of spoken narrative phrases on a simple voice recording device where the AAC user can 'tell' their story by pressing a single switch/button [10][2]. Some AAC developers have explored how narrative can be supported using templates. For example, Figure 1 shows a dynamic story template system available in a DynaVox Pageset.

While AAC can be used in many communicative contexts, most current uses of AAC focus on transactional communication. This is the sort of communication that expresses needs, wants, and information transfer, such as, "I am thirsty", or "I use a straw for drinking". Transactional communication is less dynamic, and less interesting from a natural language engineering perspective. In this paper, we instead focus on an AAC tool that facillitates a conversation about *personal narrative*. By personal narrative, we mean someone telling a story about what happened to him or her. More importantly, this type of communication (in contrast to transactional communication) has a piviotal role in cognitive and social development (as discussed below,in Section 2.4).

An approach that considers personal narrative in AAC is particularly timely given the high proportion of internet connected AAC devices [8] and recent developments in producing an open format for AAC devices [11]. These developments mean that current systems can make better use of real-world knowledge (via the internet), and have the potential to support a larger vocabulary (via open formats).

[2]The use here of low-tech is intentional. These voice-recording devices do use technology, in contrast to e.g., laminated cards. They are however not as advanced in capability as the communication aids outlined in the previous paragraph.

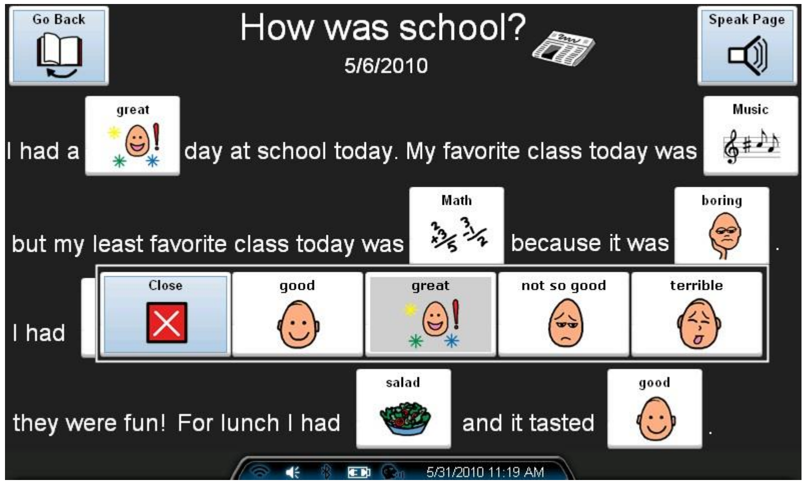


Figure 1: Example template for a child's story about their day at school used on a Dynavox system.

2.2. Potential for NLG in AAC

Natural Language Generation (nlg) systems generate texts in English and other human languages from non-linguistic input [1]. In their review of Natural Language Processing (NLP) and AAC, [12] suggest that nlg could be used to generate complete utterances from the limited input that AAC users are able to provide. For example, the Companion project [13] used nlg techniques to expand telegraphic user input, such as *"Mary go store?"*, into complete utterances, such as *"Did Mary go to the store?"*. Another approach has been for users to author utterances using Blissymbolics[3], and then to apply nlg to translate this to English and Hebrew texts [14].

The use of nlg in AAC is somewhat different from most uses of nlg. Since the goal of AAC is to help the user communicate, the nlg system must be used interactively, under the user's control; we want to assist the user in communication, not replace him or her with an nlg communicator. Also, since most human communication is social, nlg AAC systems will often need to generate texts that have the communicative goal of social interaction. Hence, nlg AAC systems are very diff t from systems that summarise

[3] BLISS is an ideographic writing system using graphical symbols representing concepts that can be combined to form words.

information in a task-oriented context, which has been the focus of most nlg research to date.

In addition to helping users interact with others, nlg techniques can also be used to support language learning, and to encourage children with disabilities. The STANDUP system for example, used nlg and computational humour techniques to allow children who use AAC devices to generate novel punning jokes [15]. This provided the children with successful experiences of controlling language, as well as giving them an opportunity to play with, and explore, new vocabulary [16]. In a small study with nine children with cerebral palsy, the children used their regular AAC tools more and also performed better on a test measuring linguistic abilities after they used STANDUP for ten weeks.

2.3. State-of-the-Art

Utterance-based AAC systems (see also [17] for an overview) are now starting to implement natural language processing, and machine learning techniques that facilitate language production on lower levels such as: improved scanning methods for letters [18]; and predicting words and utterances [19, 20]. Of particular promise are approaches that use external sources such as the internet to help develop a vocabulary for a given topic [21].

Some research projects have developed prototype systems that attempt to structure personal narrative. At their most basic, these systems provide users with a library of fi ‘conversational texts’ that can be selected and uttered. The Talk system for example, implements a retrieval system in which the user is supported to make conversational moves [22, 17]. [23] developed a system based on scripts of common interactions [24]. [25] describe a prototype utterance-based AAC system, which uses a hierarchical script (akin to the frames described by [17]) to use in particular contexts, such as going to a restaurant. Storytelling tools for AAC users have also been developed, which include ways to introduce a story, tell it at the pace required (with diversions), and give feedback to comments from listeners [10]; but all these tools are based on a library of fi texts and templates.

2.4. Personal Narrative

Personal and conversational narratives (oral stories told during interactive conversations) are crucial to social engagement. Narratives provide a means for people to relate and share experiences, develop organizational skills, work through problems, develop self image, express personality, give form and meaning to life, and allow people to be interesting entertainers [10]. Narrative skills develop experientially, with children being able to engage in storytelling even before they are verbal [26]. Early personal experience stories consist of a high point, for example, "Mummy fall!" with communication partners structuring the full story, eliciting the 'who', 'what', 'when' and 'where' [27]. However, not all experiences make good stories. An experience becomes a story if the storyteller has an emotional connection to the event [28], or if the event is unusual [29]. Parents of typically developing children encourage development of narrative skills by eliciting stories from their children [30], but the development of narrative skills is problematic for people with CCN. We recall a study where disabled children were told diff t stories more often than typically developing peers who were read the same story night after night [31]. In doing so, the children with CCN did not have the chance to learn the sequence of stories, or the structure commonly used in narrative such as beginning, middle and end. As such, initially children should use the same story template consistently until they are ready to progress to another one. It is diffi to provide access to event information which may become a story, and few AAC systems provide support for interactive story narration. However nlg gives us a possibility to change the underlying paradigm of AAC. Instead of placing the entire cognitive load on the user, AAC devices can be designed to support the retrieval of story events and the structuring of story narration for individuals with CCN.

2.5. Narrative in nlg

The computational creativity community has worked on automatic creation of fi stories for many years [32]. Most of this research has focused on high-level constructs such as plots and character development [6], although some work has also been done on lower-level linguistic issues [33]. Since the stories are fi the story generator is typically either free to make up content for the story, or to select 'interesting' content from a knowledge-base that describes the fi world. These knowledge bases are usually hand-crafted to contain interesting content, and they are not derived from real-world data.

More recently, narrative has emerged as an important research area in the nlg community; in particular data-to-text systems, which produce textual summaries of complex data sets, are more useful to recipients if their summaries are structured as narratives [34]. For data-to-text systems, the key aspects of narrative include explicitly linking related messages (typically using domain knowledge), organising texts into a narrative structure (perhaps based on a model of 'scenes'), and expressing temporal relationships in a fl t manner.

2.6. Challenges

Natural Language Generation may help structure story-telling by automatically generating utterances. However, there is a question of how to best structure these utterances into a narrative. There is also the challenge of generating the text in a way that it is suitable for a system and interface used by children with complex communication needs.

The underlying data used to generate narrative is collected as the day progresses. At any point in the day, a user may request a story is generated using the current data, even if very little has been collected.

In addition, the individual data have to be aggregated in a way that is natural and supports the sense of a story. In Section 4, we describe how sensor data can be aggregated into events (data analysis). We then evaluate these diff t methods for event generation in Section 5.

The intended user group has a very wide range of cognitive and motor skills. How we adapt to this variance in skills is described further in Section 6, including a qualitative user evaluation with two children with very diff t profi

3. The How was School Today? system

This section discusses the design of the How was School Today? system from an nlg perspective, and gives an overview of how the system evolved. The system is designed to read data from multiple sources about a child's activities and create a narrative focused on the child. The key use case is of a pre-literate child with complex communication needs telling their story to a communication partner (their parents, a member of the school staff, or peer) that they are meeting for the fi time that day.

The software generates a draft story automatically based on sensor data, and children can then edit and narrate the story using an appropriate user interface. Voice recordings can be made and embedded into the generated story, providing information that could not be inferred by the data, e.g., that a child had won a game.

How was School Today? supports children with limited memory or children who cannot communicate easily because of physical or intellectual impairment. Technology that automatically generates narrative has the potential to increase communication rates, and support dynamic personal conversation beyond what is commonly supported by state-of-the-art technology.

3.1. Feasibility study

A prototype system was developed and evaluated to assess the feasibility of the concept of automatic generation of text for narrative. While only the outcome is summarized here, the prototype is described in more detail in [35, 7]. Two children (Julie and Mary[4]) used the system for one week for a qualitative formative evaluation. Researchers supplied support during this period, primarily observing how the children used the system, and discussing improvements with teachers, therapists and parents. A third child, Eric, joined the two girls for a subsequent two-week evaluation with the next iteration of the system. In this evaluation, we asked teachers and other staff to use the system without on-site support from the researchers.

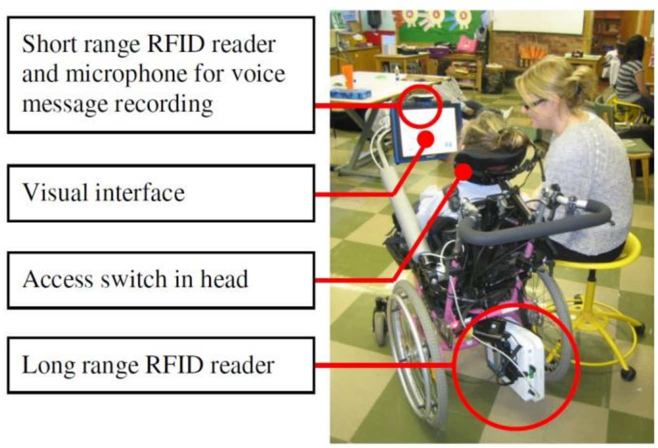


Figure 2: Participating pupil with Special Learning Assistant (SLA) using the prototype How was School Today? system: The system is mounted on the wheelchair, and the pupil has access to the system via head switch controlled row/column scanning.

The prototype system used a wheelchair-mounted Radio-Frequency Identification (RFID)[5].

[4] All examples and identifying names in this paper have been anonymised.
[5] RFID is a common technology (colloquially often called 'contactless') based on radio

The evaluation of the prototype identified three areas of improvement. Firstly, it highlighted practical issues with the hardware, such as delays caused by starting the system in the morning, and problems caused by limited battery life. A long-range RFID sensor that was used for location changes was particularly diffi to set up. Data collection required the presence of the relatively large and heavy device at all times, limiting spontaneous use.

Secondly, it highlighted issues with story generation. The system depended on the school timetable for structure, so, for example, stories could not be generated at home because there was no timetable beyond the school day. In addition, the interface did not provide access to stories from previous days.

Thirdly, it highlighted a need to fl modify the complexity of the interface: participants diff greatly with regards to how easy they found using the interface. Also, staff and parents could only make voice recordings after the fact (when they had access to the storytelling interface) rather than during data collection, again limiting spontaneous use.

The current system, described below, is based on the prototype used in the feasibility study and addresses all three issues that we identified in the original evaluation. To avoid duplication, we will only describe the current system in detail, as it encompasses most of the functionality of the original system while allowing us to emphasize the functionality that has been added.

3.2. Hardware architecture

Figure 3 describes the architecture of the system evaluated in this paper. Instead of a large heavy device, the current system uses a mobile phone as the primary data collection device, which allows participants to be more mobile, and to collect data at home as well as at the school as events unfold. We also changed the system from a stand-alone application to a client-server architecture so that stories could be told from a number of devices.

RFID tag interactions (used for locations, people and objects) as well as photos and voice recordings are collected using a dedicated mobile app, along with a suitable smartphone[6].

Staff and parents were encouraged to take photographs with the mobile phone, and were trained to record voice messages whenever something inter-

---

waves that enables the wireless reading and storing of data on tags that can be applied to, or embedded into products, people and animals.

[6]We used the Series 40 Nokia 6212 NFC SDK.

esting happened during the day [36]. The photographs were later used to identify the voice recordings in the user interface and augment the narrative.

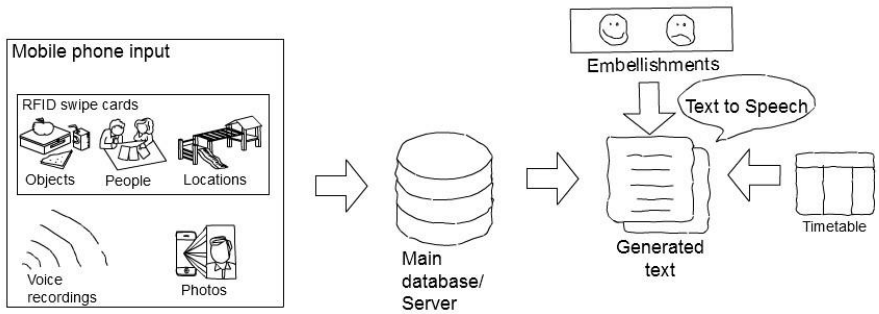


Figure 3: Types of input that can be collected by a mobile device: voice recording, RFID (objects, people and locations) and photos. This is supplemented by a timetable that teachers could modify. The data from each phone is then sent to a central database, which is called from a server, which generates the texts and allows the children to output utterance and voice recordings as well as add their own evaluations.

The transferred data is encrypted[7] and sent over a 3G network to a server where this data is written to a database. Then, when a child wants to use this data to tell the story of a given day, this data was retrieved and analysed (see also Section 4).

The generated stories are shown on a PC or a DynaVox Communication device[8], and retold using off the shelf text to speech[9] to output the nlg generated utterances together with the photos and voice recordings.

We also experimented with automated Wi-Fi location tracking[10] to replace the 'manual' RFID tracking by teachers. However, this technology is not yet sufficiently accurate to avoid occasional 'ghost' readings when, for example, moving near the edges of a room.

[7]Using the Bouncy Castle crypto API: http://www.bouncycastle.org/, retrieved November 2015

[8]DynaVox Vmax, MS Windows based communication aid that allows touch screen access and combination with other communication software, http://www.dynavoxtech.com/products/vmaxplus, retrieved November 2015

[9]Ghost! Clipboard reader, http://download.cnet.com/Ghost-Clipboard-Reader/3000-7239_4-10858927.html, retrieved November 2015

[10]Wi-Fi location tracking by Ekahau, http://www.ekahau.com/, retrieved November 2015

3.3. Domain reasoning and knowledge acquisition

Adding some interpretations of the sensor data allows us to generate a more interesting story. This requires domain knowledge and hence knowledge acquisition. For example, a logging system in a museum can utilise information about the exhibits and model the behavior of typical users (see e.g., [37, 38]) to generate relevant stories. Also, it is useful to aggregate low-level sensor information into higher-level concepts which are more meaningful to users; for example Nyx[11] aggregates sensor data about respiration into sleep information, and Fitbit[12] aggregates motion data into information about exercise.

The development process followed a user-centred design, where children and staff were involved throughout the development cycle [36]. The stories are based on data collected during the course of a day, but the generation of stories also uses some domain knowledge. For example, this version of the system can use the school timetable for a specific child. Unlike the prototype that used a single static timetable, the current system makes use of several user profi es, including individual daily timetables, and allows staff to change a timetable for any day of the week by editing a simple text fi

3.3.1. Input

Table 1 gives an example of the data that was collected during the course of a day at school. Data relating to locations, objects and people were mostly collected using passive RFID tags; collection of data through photos as 2D barcodes was also supported[13]. Broadly, the information collected fell into one of the following categories:

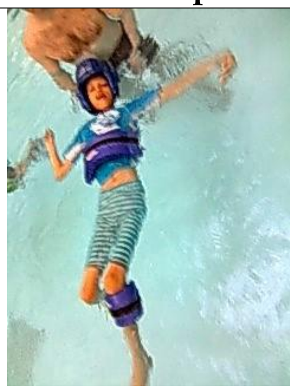
[11] http://nyxdevices.com/product/, retrieved April 2015
[12] http://www.fitbit.com/, retrieved April 2015
[13] We used the QR 2D barcode, http://www.qrcode.com/en/, retrieved November 2015

Table 1: Example data from How was School Today?

09:34, Voice Recording, A man came to talk to me in gym. I signed an important document.
09:36, Voice Recording, My dad was in school today.
09:40, Location, Gym.
09:40, Object, Skittles
09:40, Object, Baton
09:40, Person, Mrs Roberts
10:48, Voice Recording, I was doing relay racing at school and I joined in really well.
10:48, Location, Changing
10:50, Voice Recording, I did some high jump and every time I jumped I had a lie down.
11:08, Location, Classroom
11:12, Object, Blackboard
11:31, Person, Mary
11:36, Location, Tutorial Room
11:36, Object, Money
11:39, Object, Monkey Game
11:58, Location, Classroom
12:00, Location, Outside
12:08, Object, Bike
12:18, Location, Classroom
12:29, Voice Recording, I asked Jim this morning what he did for his weekend.

- Location data - each time the child changes location, in particular when they go somewhere unexpected (i.e., to a location not specified in the timetable, such as the kitchen during class).
- Object interaction - each time the child interacts with an object (e.g., a toy, a learning object, food).
- Person interaction - each time the child interacts with a person (e.g., member of staff, a friend, a visitor).
- Voice messages - staff and teachers are encouraged to record voice messages, as if the child was speaking in the fi person, talking about their experience. These can either be a single short message, or a collection of several short messages as a multi-part voice recording. Every recording can be linked to a photo to aid recall of when, and where, something occurred (see Table 2 for an example).
- Photos - staff and teachers can also save photos without voice messages. The camera feature of the phone can also be used as an alternative input of entities (i.e., objects, people, locations) through 2D barcodes.

Table 2: Example voice recording, taken during the course of the project. Note that this is a four part message, all of which are associated the photo on the left.

Associated photo	**Voice Recordings**
	Message 1: "I have just come back from swimming this morning I had good fun." Message 2: "I started off getting weights put on my legs so I could practice walking in the pool." Message 3: "Then I get the helmet on and the weights are taken of and some fl so I can swimming on my own which I like doing." Message 4: "When I was swimming so fi of all the funniest thing of the day was when (peer) came over and tried to give me a big kiss."

3.3.2. Domain knowledge

The system uses a simple domain model consisting of a relational database that holds information about RFID tags and to what they have been applied. Examples of objects (Table 3), locations (Table 4) and people (Table 5) are provided. In addition, the defi of these three entity types was supplemented by a table that stored all RFID tags defi their unique entity ID, see Table 6. This latter table affords the system the fl y to recognize these through different input methods: either as RFID tags, or (recognizing photos of) 2D barcodes.

In addition to the defi of objects and people, our system also uses a model of the child's intended activities during their school day. This takes the form of the child's school timetable (different for each day of the week), which is entered as tabulated text fi that staff can edit. Table 7 provides example information from a child's timetable for one day of the week.

The system can use this timetable to infer activity and location when no sensor data is available. If locations are swiped, these readings supersede the timetable inferences. The rationale for this is that deviations from the timetable are common and potentially represent interesting events. To this end, staff were instructed to swipe locations only when they deviated from those specified in the timetable. This reduced the amount of time staff needed to spend on data collection.

entityID	name	type	possession	article
1	skittles	PHYSIO	0	INDEFINITE
2	baton	PHYSIO	0	INDEFINITE
3	money	TEACHIN	0	INDEFINITE
4	monkey game	TOY	0	DEFINITE
5	bike	TOY	0	DEFINITE
...	...	...	...	...

Table 3: Representation of the objects used in the example data in Table 1. The type helped us identify the identify the correct verb (e.g., play or work). Possession indicated whether the object was owned by the child (e.g., my bike) and article was either definite or indefinite (e.g., the baton, or a baton).

entityID	roomname	referringnp
6	Changing	changing
7	Classroom	the classroom
8	Tutorial Room	the tutorial room
9	Outside	outside
...	...	...

Table 4: Representation of the locations used in the example data in Table 1. Here the human readable version of the referring noun phrase is tied to a more formalized version of the room name, as well as a unique label ID.

In addition, the symbols used on the narration interface (see Section 6.2.2) were personalised to each child to reflect the symbol sets already used by individual children. These were modifi and stored accordingly on their communication device to personalize the appearance of the graphical user interface with which they interacted.

3.4. Selecting what children can say

The things children can say are selected from a knowledge base created by data analysis. This module selects the fi e most interesting events to present to avoid overloading the child. If the child asks to see all the events for a given day, the system can display up to twelve events. Since we used voice

entityID	name	relationship	gender
10	Mary	ASSISTANT	FEMALE
11	Mrs Roberts	TEACHER	FEMALE
...	...	...	

Table 5: People example

entityID	taghexdata
1	7CC99DD3
6	DC8E9CD3
10	9C2F9DD3
...	...

Table 6: Labels example. This table affords the system the flexibility to recognize both RFID tags and photos of 2D barcodes.

recording based clustering, a higher number of events is unlikely given that voice recordings that happen within a reasonably large time window (e.g., 15 minutes) were aggregated into the same event. In addition, there were never more than 12 voice recordings made in our fi testing.

The fi e events are selected using a simple weighting scheme that ranks the events in order of potential 'interestingness' to the child. A larger weight is given to events with more recorded messages and interactions. Finally, the current date is included as a simple introduction to the story to orientate the child and listener.

Once the determination of events is complete, the rest of the document structuring is straightforward: the messages within an event are told in chronological order and messages are always expressed in the same way (activity description fi then interactions with people and objects). In principle, message ordering could be made more complex; we could aggregate messages describing object interactions with the activity description - ***"I played with skittles and baton in Gym with Mrs Roberts"*** for example. This facilitates an interactive conversation, giving the communication partner an opportunity to prompt for more details.

id	start	end	activity	location	teacher
1	09:00:00	12:00:00	Community Visit	Outside	Mr Jones
2	12:00:00	12:45:00	Lunch	Hall	Mr Jones
3	12:45:00	14:30:00	Class	Classroom	Mr Jones
4	14:30:00	15:00:00	Get Together	Lunch Hall	Mrs Peterson

Table 7: Timetable example with four timetabled items. It is stored as a text file which is named according to a unique child ID and day of the week.

3.5. Linguistic style

The output from the previous steps is human readable, but is not formatted as sentences or something a child would say. So, the fi nlg process converts the messages into sentences and utterances. The system resolves basic reference (e.g., use of pronouns), aggregation (e.g., *"I played with skittles and baton."*), and lexicalisation. Lexicalisation involves mapping messages to simple syntactic templates for expressing positive/negative evaluations entered by the child, and considering the right verb for a certain action (e.g., "play" when interacting with toys). To give some examples of evaluations, a positive evaluation of the teacher Mrs Roberts would be expressed as *"nice"* (e.g., *"She's nice."*), but a negative evaluation of a food object would be expressed as *"Yuck"*. Evaluations are also described further as a means of interaction in Section 6.2.2.

If the location of an event is known, this is used to describe the event (classroom – *"I was in the classroom."*), otherwise the activity (literacy – *"I was at literacy."*) is used. If no location data is given, this is inferred from the timetable. If using timetable data, activity is used before location as it is more likely to be accurate[14]. If no timetable is available a temporal marker such as *"This happened in the evening."* is used. Our microplanning is intentionally simple since it is based on a (simpler) model of speech that young and developing children would use if they were able to speak. For example, in one version of the system used in the feasibility study the microplanner dynamically determined the tense of sentences using a Reichenbach-like model

[14]We found that there were often exceptions from timetable, and that a scheduled event often took place in alternative locations.

[39]; the resulting texts were judged to be too syntactically sophisticated (especially when sentences appeared using perfect constructs), so we reverted to using simple past tense throughout.

Realisation is done using the *SimpleNLG* package [40], and the generated utterances are output using a speech synthesiser.

4. Study: Generating stories from sensor data

The prototype system segmented events according to location as indicated by a timetable. This meant that the system could not generate stories without a timetable, or if location data was missing. To improve fl xibility we developed a range of methods for segmenting the data into distinct event clusters. In this section, we give a formal specification of, and the motivation behind, each algorithm.

As the child goes about their normal school day, our system logs all the above mentioned data in the database. The input data to How was School Today? (see example in Table 1) includes information about location, interactions with objects and people, and voice messages recorded by school staff, as well as photos. The fi step is to clean up the data. For example, depending on the location sensor used, location information may be repeated (e.g., we could be told several times that the child is in the classroom), and there may also be spurious location readings (for example, an RFID location sensor was triggered when the child went by, but did not enter a room). This is done using simple heuristics which were tuned using actual data. Next, the system supplements the sensor readings using a simple knowledge base, and a child' s timetable. Notably, the timetable information specifies where the child is supposed to be at a particular time, with what teacher, and studying which subject (or performing which activity).

In order to segment the collected data into meaningful events, we fi needed to decide what constitutes a meaningful chunk by defi events. This means not only deciding where one event ends and another begins, but also which data should be omitted. Then, from the generated events, the system has to rank them to decide which are the most important.

Some of the methods we considered merely segment the data; others have a certain (characteristic) degree of lossiness. We compared three algorithms in an experiment; these were location-based clustering (similar to the timetable driven stories used in the prototype system), recording-based clustering (used in the current/improved system), and time-based clustering (for comparison). For all three algorithms, an event is defi the same way, with a location and any associated people, objects, and multi-media. For reference, Table 8 (on the next page) shows how each algorithm fi and clusters the example data given in Table 1.

Each of these methods is motivated by a diff t model of how children structure a narrative about an important event in their day. For more information about the three algorithms, see [41]. Note that diff t algorithms can result in noticeably diff t stories, e.g., the location-based algorithm results in many short events, and the voice recording-based approach in fewer but longer events.

4.1. Location based clustering (LOC)

This approach considers that a change of location is likely to denote a major change of context, and can be used as an event boundary. Once a new location is identified, the additional sensor data (who was there, what objects were interacted with, any voice recordings) is tied to the event. If there is no data associated to a particular location change, the event is deleted. The underlying justification for this fi is that uninteresting narratives are likely to result from rooms that have been passed without anything occurring in them. The location based clustering algorithm considers all non-location datatypes to be equally important. In practice, this algorithm could set an upper bound for time, creating a new event after a given number of minutes without a location change to be able to generate stories for children who conduct much of their activity in the same location. An example cluster generated from our data is:

11:36:00, Location, Tutorial Room
11:36:00, Object, Money
11:39:00, Object, Monkey Game

When converted into English text, the above cluster gives the story: *I played with Money and Monkey Game. This happened at a Tutorial Room.*

Location Cluster	Resultant Story
09:34 Voice Recording, A man came to talk to me in gym. I signed an important document. 09:36 Voice Recording, My dad was in school today!	A man came to talk to me in gym. I signed an important document. My dad was in school today.
09:40 Location, Gym. 09:40 Object, Skittles 09:40 Object, Baton 09:40 Person, Mrs Roberts 10:48 Voice Recording, I was doing relay racing at school and I joined in really well.	I played with Skittles and Baton. Mrs Roberts was there. I was doing relay racing at school and I joined in really well. This happened at the Gym.
10:48 Location, Changing 10:50 Voice Recording, I did some high jump and every time I jumped I had a lie down on the mat.	I did some high jump and every time I jumped I had a lie down on the mat. This happened at the Changing Room.
11:08 Location, Classroom 11:12 Object, Blackboard 11:31 Person, Mary	I worked with the Blackboard. Mary was there. This happened at my home classroom.
11:36 Location, Tutorial Room 11:36 Object, Money 11:39 Object, Monkey Game	I played with Money and Monkey Game. This happened at a Tutorial Room.
12:00 Location, Outside 12:08 Object, Bike	I played with the Indoor Bike. This happened outside.
12:18 Location, Classroom 12:29 Voice Recording, I asked Jim this morning what he did for his weekend.	I asked Jim this morning what he did for his weekend. This happened at my home classroom.

Voice Recording Cluster	Resultant Story
09:34 Voice Recording, A man came to talk to me in gym. I signed an important document. 09:36 Voice Recording, My dad was in school today! 09:40 Location, Gym. 09:40 Object, Skittles 09:40 Object, Baton 09:40 Person, Mrs Roberts	A man came to talk to me in gym. I signed an important document. This happened at Gym. Mrs Roberts was there. I played with Skittles and Baton. My dad was in school today.
10:48 Voice Recording, I was doing relay racing at school and I joined in really well. 10:48 Location, Changing 10:50 Voice Recording, I did some high jump and every time I jumped I had a lie down on the mat.	I was doing relay racing a school and I joined in really well. I did some high jump and every time I jumped I had a lie down on the mat. This happened at the Changing Room.
12:18 Location, Classroom 12:29 Voice Recording, I asked Jim this morning what he did for his weekend.	I asked Jim this morning what he did for his weekend. This happened at my home classroom.

Table 8: Stories based on sensor data, as generated by two of the three proposed algorithms. These stories are representative of the style of stories generated by the algorithms. For example, the location-based algorithm results in many short events and the voice recording-based approach in fewer but longer events.

Algorithm 1: Location based clustering

```
function locationCluster(E){
   C ← A ← ∅
   for i ∈ E (In ascending time order)
      if type(i) = location
         C ← addset(A, C)
         A ← {i}
      A ← A ∪ {i}
   return C
}

function addSet(A, C){
   if |A| ≥ 1 // (Remove "I entered the room" data)
      Construct a copy of A with the label A_|C|
      C ← C ∪ {A_|C|}
   return C
}
```

4.2. Voice recording based clustering (VR)

This algorithm takes the set of voice recordings as a starting point for a set of events, and clusters the other types of input around it. It is based on the idea that a facilitator (such as a teacher supporting a child with complex communication needs) will only make a voice recording or similar detailed note if some reportable event has occurred. Voice recordings within a certain time window (e.g., 15 minutes) are merged together to account for facilitators recording multi-sentence messages. Similarly, we considered that any sensor data that was recorded after a certain time period (e.g., 60 minutes) was no longer relevant, and could not be included in the same event. This algorithm is lossy: potentially removing 'uninteresting' episodes, but possibly also removing too much information, the previous example event (for location based clustering) would not be available to the user of the system (e.g., a child with complex communication needs).

Note that this algorithm could just as easily use another type of sensor reading to indicate that an event is reportable, and to cluster around those "important" sensor readings. For example, if the system detects that RFID cards for two or more people have been scanned, other sensor readings happening around the time of this "event" could make a story in the same way. One of the clusters generated from our data is:

09:40:00, Location, Gym.
09:40:00, Object, Skittles
09:40:00, Object, Baton
09:40:00, Person, Mrs Roberts
10:48:00, Voice Recording, I was doing relay racing at school and I joined in really well.

When converted into English, the above cluster gives the story: *I was doing relay racing at school and I joined in really well. I went to the gym today, Mrs Roberts was there, I played with skittles and the baton.*

Algorithm 2: Voice recording based clustering of sensor data

---

function $VRCluster(E)\{$
  $C \leftarrow \emptyset$
  $currentPeople \leftarrow currentObjs \leftarrow currentLoc \leftarrow \emptyset$
  for $i \in E$ (In ascending time order)
    if $type(i) = voicerecording$
      construct $A_{|C|}$ as $\{i\} \cup currentPpl \cup currentObjs \cup currentLoc$
      $C \leftarrow C \cup A_{|C|}$
    else if $type(i) = location$
      $currentPpl \leftarrow currentObj \leftarrow \emptyset$
      $currentLoc \leftarrow \{i\}$
    else if $type(i) = person$
      $currentPpl \leftarrow currentPpl \cup \{i\}$
    else if $type(i) = object$
      $currentObjs \leftarrow currentObjs \cup \{i\}$
  return C
$\}$

4.3. Time based clustering (T)

Algorithm 3: Time based clustering

---

function *timeCluster*(E){
  $C \leftarrow \{\{i_1\}, \{i_2\}, \ldots, \{i_{|E|}\}\}$
    (Each element starts in its own 'cluster')
  while $|C| > numberOfEventsRequired$
    Let X be the set of i, j pairs such that $i, j \in E, \{i, j\} \not\subset A$, for any $A \in C$.
    Choose $(i, j) \in X$ such that $time(i) - time(j)$ is smallest and let $a = i, b = j$
    Then $a \in A, b \in B, A, B \in C$
    $N \leftarrow A \cup B$
    $C \leftarrow C ¥ \{A, B\}$
    addSet(N, C)
  return C
}

function *addSet*(A, C){
  if $A \neq \emptyset$
    Construct a copy of A with the label $A_{|C|}$
    $C \leftarrow C \cup \{A_{|C|}\}$
  return C
}

The time based clustering algorithm creates a greedy/hierarchical clustering based on temporal proximity. Data elements are grouped according to the temporal proximity of the data; so things that happened around the same time are likely to belong to the same event. This algorithm fi considers each element to be its own unique cluster, and then seeks the two elements that are closest together in time outside of the cluster. Once found, the two relevant clusters are merged into one cluster. This process continues until a predetermined number of clusters remain. The time based algorithm is able to implicitly follow the natural boundaries of the events of the day, allowing smaller events to be combined into larger ones if necessary, and spreading out the stories evenly.

This method does not loose any data, and the number of events that it produces can be specified in advance. However, deciding the optimal number of clusters is a parameter that may need to be adjusted for a given dataset. One of the clusters generated from our data is:

11:31:00, Person, Mary
11:36:00, Location, Tutorial Room
11:36:00, Object, Money
11:39:00, Object, Monkey Game

When converted into English, this gives the story: *Mary was there. I went to a Tutorial Room. I played with Money and the Monkey Game.*

It is notable that this story is only larger by one fact than that produced by the location based clustering method; however, in this case, what had actually happened is that Mary had escorted the child to the tutorial room, and played the money and monkey game with the child. So, using the timing of events has helped us to produce a more realistic story.

5. Experiment

This section describes a study where twenty-six parents evaluated the three algorithms for clustering described in Section 4 on their degree of realism, completeness, readability, and how much they made sense to the reader. The participants were parents recruited outwith the special needs schools, and the stories were based on real data from a special needs school (more details below).

5.1. Participants

Twenty-six parents and grandparents of children aged 7 to 15 took part in the study. These were recruited outwith the participating special needs schools. The majority (69.2%) of these parents were aged 35-45, with some aged 25-35 (11.5%), and some over 45 (19.2%). 18 of these participants were female, 8 were male. All participants were either native English speakers (92.3%) or fl t in English (7.7%). Three of these parents had children with special needs. The number of children in our participants' household ranged from one, to more than four, with most parents having one (23.1%), two (24.6%) or three (38.5%) children.

Table 9: Count of different types of data (words, sentences, objects, people, voice recordings, locations, sentences per event, events) retained for the different algorithms (T= time based, LOC = location, VR = voice recording) in datasets D1-D4

		words	sents.	obj.	ppl.	voice rec.	locs	sents./ event	events
D1	T	154	24	6	2	5	11	5.75	4
	LOC	132	19	6	2	5	7	2.86	8
	VR	86	11	2	1	5	3	3.67	3
D2	T	217	29	10	7	7	10	5.8	4
	LOC	195	25	10	7	7	7	3.7	7
	VR	125	12	2	0	7	3	4.0	3
D3	T	216	30	2	10	7	14	5.0	4
	LOC	188	25	2	10	7	9	3.0	9
	VR	144	16	1	4	7	5	3.2	5
D4	T	225	34	1	16	10	13	6.8	5
	LOC	186	26	1	16	10	7	3.7	7
	VR	159	20	1	4	10	5	3.8	5

5.2. Procedure

Using questionnaires[15], participants evaluated the three algorithms: time based, location based, and voice recording based clustering (T, LOC, and VR) for four diff t datasets (D1-D4). The order of the datasets and the order of the algorithms were both randomised. For each data set, participants were asked to rank the three versions of the same story on four dimensions: how realistic they were; how much they made sense; how complete they were; and how easy they were to read. In some cases participants were not able to decide between stories, and were asked to rate those stories equally.

5.3. Materials

The materials consisted of four datasets D1-D4. Suitable data was collected from two special needs schools, each providing two datasets, spanning diff t days and children. For datasets D1 and D2, this data was manually collected by shadowing a child for a day, and collecting object and person

[15] http://joereddington.com/csr/hosted/schoolDomainQuestionnaire.pdf, posted January 2012

Table 10: Maximum possible count of words, sentences, objects, people, voice recordings and locations for the used dataset texts (D1-D4).

	# words	# sentences	# objects	# people	# voice rec.	# locations
D1	154	24	6	2	5	11
D2	217	29	10	7	7	10
D3	216	30	2	10	7	47
D4	225	34	1	16	10	40

interactions, location changes and voice recordings (by means of a custom mobile phone application).

Datasets D3 and D4 are based on automatically collected data that was later anonymised. This data was similarly coded, and only deviated with regard to the number of location changes. The larger number is due to the true nature of movement and a sensor problem, which often registered many passages from one room to another within the same minute.

A copy of each data set was split into event boundaries by one of the clustering algorithms, and the resulting clusters were converted to simple natural language using a few simple templates such as ***"This happened at [locationname]"***, and ***"I played with [toy object]"***. Finally, the generated texts were anonymised to preserve the privacy of the participants.

Tables 9 and 10 summarise some properties of these datasets. It is clear that the time based algorithm (T) is essentially loss-less, save for fi ring out incorrect location changes. The stories generated based on voice recordings (VR) are generally half the length of the time based approach. Stories based on location changes (LOC) are generally segmented into more, but shorter, events.

5.4. Results

Stories based on VR performed better than expected given that they represent around half of the information used in the stories generated using T. In many cases they were preferred by a majority of participants, in particular for ease of reading ("Easy"), and how much sense they made ("Sense").

However, participants were consistently able to tell that the stories based on T were the *most complete* (“Complete”). T also performed well in terms of ease of reading and how much sense it was thought to make. Table 11 shows a summary of the results. The distribution was relatively equal between the three clustering methods for the dimension “Likely” , and the results are included for completeness.

Table 11 also shows that the results are largely consistent across datasets for the Easy and Sense dimensions. In some cases, this diff is small, and in others it is statistically significant using a Chi-square test (omitting ties). A deviation is seen for the fi dataset, D1, where T performs well on all three dimensions. When examining the properties of the D1 dataset in Tables 9 and 10, it can be observed that this dataset is similar to the other three datasets in terms of the percentage and type of information that is lost. However, the original data for this dataset has the fewest items, leading to an even shorter fi story, which may be a contributing factor to the observed results.

Further analysis was performed to see if there were any properties of D2 and D3 compared to D4 that might explain why the latter dataset did not have significant diff between conditions. The most notable diff are the relative length and number of voice recordings. The length is shorter for D2 and D3, both of which have 7 voice recordings rather than the 10 used in D4. This may suggest that there is an optimal range for the length of stories: longer than D1, and shorter than D4.

All stories are highly readable as measured by the Flesch metric [42] with scores over 80 (none of the stories contained passive sentences), out of a possible 100, with high scores signifying the higher readability. This implies that a preference for one method over another in the survey was not due to one of the clustering methods creating stories that are diffi to read.

In summary, we found that the voice-based clustering resulted in the most realistic (for the age group), and most easy to read stories. This is why this is the method we implmented in the trialled system described in later sections of this paper.

5.5. Discussion

Participants were also encouraged to leave comments about the stories they read, and this section discusses a number of reoccurring issues that came up in the written comments. Some of the participants felt that there were too many facts, and too few emotive comments in the stories. Future

Table 11: Distribution percentages (and absolute number in parenthesis) for the winning strategy on the dimensions "Easy", "Complete", "Sense", and "Likely" for the three al- gorithms across four datasets. The highest scoring algorithm is highlighted in each cell, but is not always significant. * denotes significance at $p <= 0.05$, and ** significance $p <= 0.01$ in Chi-square tests. Distribution of tied results is not included in the table, but is in the calculation.

		"Easy"	"Complete"	"Sense"	"Likely"
D1	T	**42.3** (11)	**61.5** (16) **	**38.5** (10)	26.9 (7)
	LOC	26.9 (7)	26.9 (7)	26.9 (7)	**30.8** (8)
	VR	26.9 (7)	7.7 (2)	23.1 (6)	26.9 (7)
D2	T	26.9 (7)	**46.2** (12)	15.4 (4)	26.9 (7)
	LOC	19.2 (5)	23.1 (6)	7.7 (2)	**34.6** (9)
	VR	**42.3** (11) *	26.9 (7)	**61.5** (16) **	26.9 (7)
D3	T	30.8 (8)	**50.0** (13) *	26.9 (7)	15.4 (4)
	LOC	11.5 (3)	23.1 (6)	26.9 (7)	26.9 (7)
	VR	**50.0** (13) *	11.5 (3)	**38.5** (10) *	**30.8** (8)
D4	T	30.8 (8)	**34.6** (9)	19.2 (5)	15.4 (4)
	LOC	15.4 (4)	26.9 (7)	23.1 (6)	26.9 (7)
	VR	**46.2** (12)	23.1 (7)	**46.2** (12)	**34.6** (9)
All	T	32.7 (34)	**48.0** (50) **	20.0 (26)	21.2 (22)
	LOC	18.2 (19)	26.0 (27)	21.2 (22)	26.9 (28)
	VR	**41.3** (43) *	18.2 (19)	**42.3** (44) *	**29.8** (31)

iterations of the system will include more in the way of emotive annotation such as ***"She was nice."*** (positive evaluation of person statement) or ***"Lunch was yuck."*** (negative evaluation of food statement) as in [44].

Moreover, parents found that there was a divergence between what they might want to hear from their children and what the children might be likely to say. Many parents offered the opinion that the most likely reply by a child to the question "How was school?" would be the word "fine". Similarly, many parents learned about their children's day at school from other parents. In this manner, two parents can exchange the stories they heard from their child to form a more complete story of their day. A story generation system in the domain of AAC needs to consider that this may also be the depth of description a child might want to go into. Naturally, this is an area where the children's privacy and integrity may be at stake and any advancement in this area must be followed with great caution ([45] discusses some of the privacy issues in utterance-based AAC).

One of our participants suggested that information such as wheelchair adjustments might be an example of the type of information that might be useful for a parent to know, but rather unlikely to be something the child would say. Other parents thought that their children would be likely to talk about their peers at school.

Parents also noted that the language was more American English than U.K. (where the study took place), and that it was not always age appropriate. When generating natural language, it is important to consider regional diff and indeed using NLG techniques allows for just this type of fl ibility in terms of lexical decisions (word choice). Since the children of these parents ranged from 7 to 15, it is understandable that the texts might not have been age appropriate for all the ages. Possibly this is due to the complexity of language introduced via voice recordings. While staff may record voice recordings in fi person, it may not be phrased in the way a child would.

6. Qualitative Evaluation

In the previous section, we found that clustering by voice recording resulted in stories that were perceived by parents to be easy to understand, and to make sense. The improved system therefore uses this clustering method. This section presents a qualitative evaluation of the improved How was School Today? system in the school environment.

The evaluation of the system included two sub-components: a) mobile phone, with a sensor for RFID tags, camera for 2D barcode recognition, and the ability to record messages; and b) the main storytelling device running on the child' s Voice Output Communication Aid (VOCA). This system was developed with the school (a diff t one from the feasibility study) by two children over a six months period. The fi four weeks served as more formal trial of system.

6.1. Participants

In collaboration with staff at the school, eight possible participants were initially identified on the basis of their communication and/or intellectual impairments; some of the children exhibited a desire to share their experiences, for others it was felt by staff that the children might benefit from participation. A bigger pool of potential participants allowed us to prepare for drop-outs due to illness and other issues (such as key staff leaving), which are common when working with individuals with severe disabilities.

After drop-outs, we continued the evaluation with two children with very diverse profi in particular in terms of mobility and age. The participants, their classes, teachers, and Special Learning Assistants (SLAs) took part in this evaluation. Other members of staff included kitchen staff, teachers for specialized subjects (e.g., Home Economics) and other staff who were likely to be in contact with the participants for day to day usage. Data collection for both participants (RFID, barcodes, voice messages, and photographs) was undertaken by staff. Names are changed to preserve anonymity.

Peter. Peter has athetoid cerebral palsy, uses a wheelchair and is not independently mobile (he depends on others to push his chair). He has very little functional speech, having a few recognisable words. He uses gestures and head pointing in his environment for basic communication as well as an E-Tran[16] folder for aided communication. Peter can use switches to access a computer using his hands for selection, but this access method is inefficient due to Peter' s limited motor control.

Martin. The second child, Martin, has a chromosomal disorder, is ambulatory and independently mobile in the school environment. He has very little functional speech, using some single words. He uses Makaton[17] gestures, and points at things in his environment. He also uses some graphic supports (photos and symbols) for communication.

6.2. Set-up

This section describes the baseline and the set-up used for the two children. For a more detailed description of the setup, see also [36].

[16] Eye-Transfer (E-Tran) systems are "low tech" AAC devices using eye pointing for spelling or access to words and phrases. Peter first selects one of 4 or 6 color coded groups, and then selects the item with a second gaze at the appropriate color.

[17] Makaton uses signs together with symbols and speech to support communication for individuals with communication and/or learning difficulties. http://www.makaton.org/, retrieved November 2015

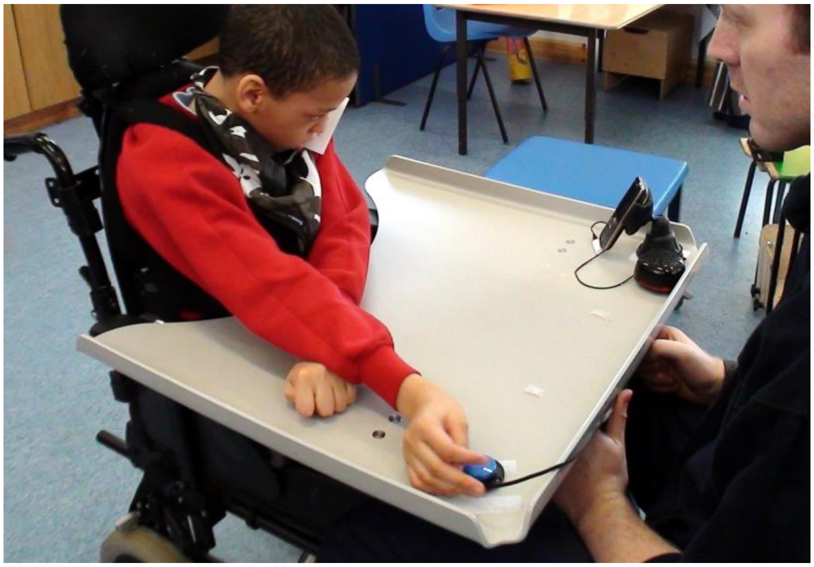


Figure 4: Peter using a switch with his right hand to access the mobile phone for sequential playback of recorded voice messages.

6.2.1. Baseline

Prior to introducing the system, there was no comparable communication aid for the children. Therefore, the two children were equipped with Step-by-step[18] devices. These allow staff and parents to record sequenced voice messages, which were accessible to the children through a single switch.

In addition to providing a baseline, this prepared both the participants and their environment for successful use of voice recordings for personal narrative. The successful use of a narrative tool that uses voice recordings depends on the skill of the people who record the voice messages and other data onto the system in order to successfully support an interactive narrative [46].

Objects of interest for the two participants were identified and RFID tags were attached to them (such as a game and coins for number/money teaching). The HEX identification codes of the RFID tags were stored in the system database. A4 sized signs with both, a) a hidden RFID tag, and b) a large (15cm x15cm) barcode, were attached to all rooms accessible to the

[18]See https://www.ablenetinc.com/little-step-by-step-with-levels, retrieved November 2015

participants, including doors leading outside (for activities such as garden work or community trips with the bus).

6.2.2. Intervention: narrative with How was School Today?

The How was School Today? system supports the children that use it by structuring the storytelling process. It recalls the relevant data as well as tackling the task of composing correct sentences from this data, thereby reducing the cognitive load on the child. Once the text has been generated, other key functions of the system are to support the telling, and the editing of stories (hide events or show previously hidden events).

Also, the two children we worked with diff in terms of cognitive, motor, and social ability. The need to support a wide range of abilities was clear even in our original feasibility study, where we found that our interface worked well for one child, but not another [35]. The amount of functionality available to the child compared to that available to the support worker/carer therefore diff The amount of control available to the children also diff according to their abilities.

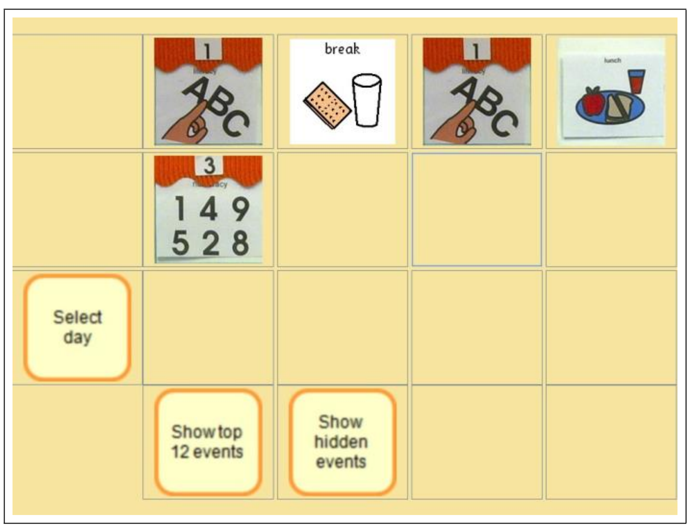


Figure 5: An interface for children at 'Stage 3', who are able to select a day they want to talk about. This day has no data, but shows the first five events that are on the timetable for that day.

To see how this worked in practice, let us fi look at narration. First, an event is selected. This event, like every event, can consist of multiple

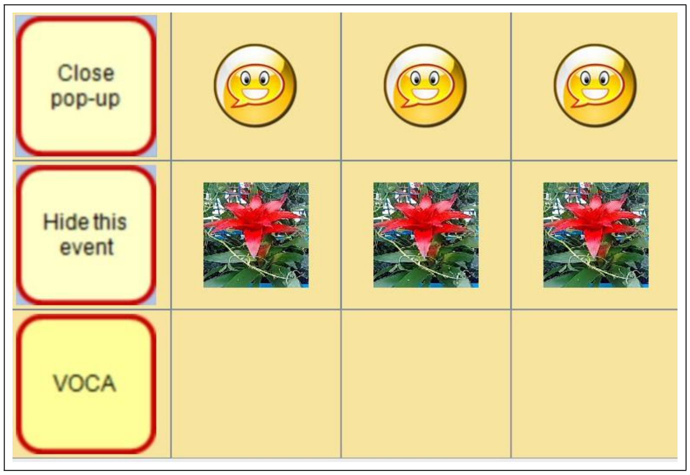


Figure 6: Screenshot for a single event with 3 generated utterances and 3 audio recordings. The three recordings are actually part of a single larger message and were taken at the same time as a single photo (a plant which was drawn during art class). Each recording is a single phrase in first person, but relates to the same event and is therefore tied to the same photo.

messages. These messages are either a generated phrase or a voice recording (see Figures 5 and 6). For children with severe motor and cognitive impairments, events and messages can only be told in a sequential manner, while children with less complex impairments can select the events and messages in the order they see fi While the system in the feasibility study was limited to sequential narration, the current system also supports narration in any order and three stages of abilities:

- **Stage 1**: Children with very limited skills, who in particular may not be able to do much more than press a 'Next' button on an interface. Teachers will edit content and select an event, that the child can sequentially step through. This stage was used with Martin.
- **Stage 2**: This corresponds to children who can select both the order of events and messages as well as add evaluations (e.g., 'She is nice'). This stage was used by Peter.
- **Stage 3**: This is for more able children, and brings in the ability to tell stories about previous days as well.

Depending on the stage of the system used, the fl w of storytelling can be controlled on a number of other levels such as selecting what day to talk about, what event during that day, and what piece of information (or event related message) to share about that day.

Children at Stages 2 and 3 can also hide events, or show previously hidden events. This gives children the ability to completely omit parts of the story they do not want to talk about, or that may be incorrect.

Each message in a story is either an utterance (based on sensor data) or a voice recording. Each message also has two evaluations with it: a happy, and a sad face. These evaluations automatically generate an utterance about the message such as ***"I don't like playing with the bike."***, if the utterance was about the bike and they selected the sad face. In Table 1 (p. 13), this reflects a sensor reading at 12.08 for a bike, and the fact that this is a toy in the knowledge base means we know that the appropriate verb is 'play'.

6.3. Trial period

During the six months the researchers visited the school regularly to a) support the data gathering process, b) access data for storytelling using the mobile phone and c) establish the user interface access on a VOCA. All participants (including parents and staff) were encouraged to use the mobile phone for data collection and use the phone software's playback function to share stories based on the voice recordings, both in school and at home. Peter played back messages that were pre-selected by staff. The phone was placed in a cradle on his lap tray (see example message in Table 2), and Peter played back the messages sequentially using a switch that was connected to the phone. Martin was able to play back messages by pressing the center navigation button on the phone, but had little concept of navigating to the messages by himself and needed support from staff to select stories he might have wanted to tell.

The user interface on the VOCA was developed parallel to the introduction and use of the mobile phone. Initial testing of the interface was performed using a Wizard of Oz [47] setup, with researchers manually adding utterances based on the collected data. These utterances were retrieved from the system back-end, which was completed before the fi user interface was developed. This approach had several advantages. Firstly, this interface was used to familiarise the children with the system. It gave the children an opportunity to get familiar with the concept of a system that structured personal narrative, and to practice using it. Secondly, it helped us improve the design of the user interface before implementing the fi version of the system.

The Wizard of Oz setup highlighted a few practical areas for improvement, including the importance of using familiar images for each participant. We worked closely with teachers and SLAs to make sure that the images used were based on what the children had previously used in their literacy and English classes. We also found that the children diff red in terms of their preferred input mechanisms, and degree of control they were comfortable with (see also Section 6.2.2).

Figure 7: Wizard of Oz VOCA Interface using The Grid 2 software to display and access recordings and automatically generated messages based on collected data. This screen show the messages for a specific event (green fields, from top left): location message ("I was in the gallery"); people interaction message ("Peter, Paul and Mary were there"); object interaction message ("I used my reading book"'); and four voice messages, labeled with photographs taken at the time of recording. The page also contains evaluation messages (such as "I liked it"), buttons for navigation ('back to day overview', top left) and editing ('hide this event', middle left), and a link to the child's regular AAC method (bottom left).

Figure 7 shows the Wizard of Oz interface for Peter. During the evaluation Peter was equipped with a new VOCA with eye gaze access (Tobii

C12[19]), which he started using to access the stories. However, due to the novelty of the device use, he still had considerable diffi in accessing the stored messages effectively. Martin tried a similar setup on a touchscreen VOCA (DynaVox Vmax[20]), but had diffi navigating and preferred using a single switch for accessing stories as sequentially played back messages to selecting individual messages via the touchscreen.

In the last four weeks of the evaluation, the Wizard of Oz interface was replaced with a fi version of the system. Figures 5 and 6 demonstrate the user interface used in the fi version.

6.4. Outcomes

The evaluation in the second school focused on two questions:

- Independent use: Could the system be used independently in the school?
- Support for narrative: How would personal narrative change with using the system in comparison to previous methods of personal narrative support?

6.4.1. Independent use

Conversations of the participants using diff t means of support were observed and feedback from all involved individuals gathered. Although the mobile phone could be used independently, the VOCA based narrative software did not reach a level that it could be used independently by staff or parents. Instead the researcher set up the system in the school for a small number of conversations of the children with parents and staff to collect impressions on its use.

Questionnaires were sent out to the teachers and SLA's. The majority (13 out of 14 questionnaires) used the phone to record messages, take photographs or to detect RFID and barcodes for data input. Half of them (7 out of 14) found the phone easy to use, three stated from "not easy" to "very diffi with four giving no feedback on diffi of phone use. The class teachers of the children stated greatly improved communication with home, and a considerable impact on the childrens' enthusiasm to share stories (the

[19]See http://web.archive.org/web/20130104211912/http://www.inclusive.co.uk/tobii-c12-p2113, retrieved November 2015
[20]See http://uk.dynavoxtech.com/products/vmaxplus/, retrieved November 2015

children were reported to immediately want to use the phone when coming into class to share stories from home, and actively seeking contact to staff outside the class to tell them their story).

Parents stated that they did not fi using the mobile phone prototype diffi and one family used it on holiday. In particular, they liked the ability to link photographic images to the voice recordings. In general, they confirmed their interest in using the system to hear about their child' s expe- riences and activities at school.

6.4.2. Support for narrative

Feedback was collected using questionnaires and semi-structured interviews from all participants (e.g., the children, parents, and teachers). Additionally, informal feedback from both Peter and Martin was collected during the evaluation. They both appeared to enjoy taking part throughout. A semi-structured interview was conducted with Peter to supplement his feedback during the evaluation period. Although Peter had never mentioned anything during the evaluation, in this interview he stated that he did not enjoy testing the mobile phone prototype. However, he stated also that using the mobile phone (with switch access for sequential message access) was easy. When asked why he did not enjoy testing the mobile phone prototype, and if it was due to the wrong stories on the phone he agreed. He had just recently been equipped with a new eye gaze VOCA, and when this was brought into the conversation Peter stated that he preferred using the prototype on his eye gaze VOCA because he had more control over what he could say.

On some days there were few, or even zero, voice recordings. The used method only included sensor data that was clustered around voice recordings, which meant that those days have very little narrative content. However, even on the days where there were no voice recordings at all, the children were able to mention events on their timetable as well as evaluations (e.g., "it was fun").

6.5. Example of Narrative Interactions

The system allowed access to stories in two ways. Voice recordings and the associated photographs could be accessed directly on the mobile phone and played back in sequential order. Martin could sequentially play the next recording by using a switch that accessed the playback function on the phone, as shown in Figure 4.

Narratives, including voice messages, generated utterances and photographs could be accessed via a graphical user interface similar to one used in the initial feasibility study. First, an event was selected from a screen such as Figure 5. Depending on the child' s ability, this would be done by the teacher (in Martin' s case), or by the child (in Peter' s case). Figure 6 demonstrates the view of the system once an event was selected. Martin used a similar sequential navigation as with the phone, where the switch allowed him to play the next utterance or voice recording. Peter could select events using eye-gaze, as well as select individual automatically generated utterances (these are the speech bubbles in Figure 6), as well as voice recordings (in Figure 6 as photographs).

7. Next steps

While How was School Today? has been a large step forward in supporting personal communication, there are a number of issues we plan to address in the long term. These include expanding the domain, supporting children with diff t levels and types of impairments, narrative across the lifespan, true dialogue narration, pragmatics of interacting with others, and security and privacy issues.

7.1. Expansion of domain

The system is currently only designed for use in a school, while there are a wide range of other contexts for it. We also believe that this work can benefit research outside of AAC such as life-logging, e.g., [48, 49, 50, 51, 52], where a lot of personal data is available, but not aggregated into a human readable narrative. We have experimented with using 2D barcodes as well as RFID to reference database entries, but in an open domain this means there needs to be a mechanism for defi new objects and people (or anything else that can be talked about). Mobile phone based 2D bar-code readers are also being used to share stories [43]. In addition, personal stories are already being recorded online and on a large scale, on social media such as Twitter [48]. However, these are not yet structured narratives. Further studies are required to examine what happens when trying to generate a narrative in a less controlled environment, where also additional location readings from GPS could be used.

7.2. Supporting children with different levels and types of impairments

As previously mentioned, a key issue in AAC is the diversity of AAC users. How was School Today? already considers some diff in terms of motor and cognitive ability by developing three stages of a modular interface, and being able to use diff t input devices. How was School Today? does not currently consider the pragmatics of conversation, such as including support for turn-taking or shifts in topic (although it supports talking about multiple events or topics). This may be of particular use to users on the autism spectrum, and in particular those with Asperger syndrome who have a strong command of language but could use support in responding to partners by varying the content of their dialogue.

7.3. Narrative across the lifespan

In [53] we expressed the vision of a tool that is able to support children over time, as their abilities grow and as their experiences accumulate. How was School Today? also intends to be a repository of a child' s personal stories. While the current system is a step in this direction, further support can be made to include a wider range of abilities. In particular, support for retelling favorite stories, and stories further in the past is something we would like to see in future AAC devices. Also, the language the system uses is most suitable for a younger age group. However, as the children grow and develop linguistically, the nlg should be able to adapt word choice and grammatical structure to the current and progressing level of the children.

7.4. True dialogue narration

Our current system incorporates a simple model of a conversation, where children are restricted at any point to choosing from a small number of options. The child chooses a day or event to talk about, and then goes through the sequence of messages associated with that event. The child has the freedom to switch to a diff t event, hence controlling the conversation, and to add annotations/evaluations (for example "it was fun!"). In our evaluation with Peter we found that it was important for him to be able to control the selection of events: it frustrated him when he found that the 'wrong' stories were being retrieved on the phone. This suggest that it is important to include the functionality of selecting events (in contrast to e.g., strict sequential narration) on the VOCA. The ability to choose day and event is adequate in many cases, but in the long term we would like to support more complex conversations; for example interrupting a discussion about today' s events to talk about a particular teacher instead of an event. We would also like children to easily be able to add conversational phrases, such as "Guess what happened..." .

7.5. Security and privacy issues

With such a range of personal information being collected, there is a clear and present need for data protection policies, and for systems to be designed in such a way that reflects human notions of privacy [45]. Future AAC devices with features derived from How was School Today? will combine features of communication aid, diary, and identity. In this 'space' the value of data is socially constructed, and that value changes with use from being a set of needs-based utterances to being an intimate part of that individual's identity.

Also, as with all automatically generated content (and in particular for a pre-literate user), the user does not know what content has arrived on their device before they use it. So the user might be even more surprised than their parents that they are telling the story "I went into the girl's changing room", rather than "I went to the gym".

As discussed in [41], there is trade off with systems that produce automatically generated content. The more of a user's privacy they sacrifice, the more accurate and worthwhile their generated utterances can be. It is critically important that support staff and users are aware of the trade off that they are making.

8. Conclusion

In this paper, we have described a model for building an NLG system for personal story telling. We present an example of a system trialled in special needs schools, and the challenges involved with building and evaluating it 'in the wild'. Also, we found that while voice recordings appear to be a good way to structure raw data for personal story generation, they are also very lossy. In a real school setting, some days there were few, or even zero, voice recordings. This means that some collected sensor data, that could have been used to generate stories, is not used to generate the personal narrative. We note however, that even in the worst case scenarios (i.e., there are no voice recordings or no data) children could always mention events in their timetable as well as supply positive and negative evaluations. An improvement would be to use a combination of clustering methods depending on how many events are generated using a particular method. Another solution may be clustering around an importance indicator other than voice recording, preferably one that can be automatically generated rather than relying on human input. For example, an 'important event' could be inferred from the fact that RFID cards for two or more people have been scanned. The system could cluster the sensor readings around the time that this detected 'important event' occurred.

The class teachers of the participants stated that there was greatly im- proved communication with home, and a considerable impact on the partici- pants' enthusiasm to share stories: participants were reported as immediately wanting to use the phone when coming into class to share stories from home, and actively seeking contact with staff outside the class to tell them their story. Parents stated that they did not fi using the mobile phone soft- ware diffi and one family used it on holiday. In general, they confirmed their interest in using the system to hear about their child' s experiences and activities at school.

While the How was School Today? system is a step forward for the children, parents and staff involved in this study, there are a number of open avenues for future research. These include improved support for dialog across diff t ages, improving privacy of personal stories, as well as pragmatics in conversation. In our future work, we plan to investigate the role of the system in education over time: as part of speech and language therapist practices, and the development of imagination and cognition of participating children.

Acknowledgements

The authors thank the children, their parents and school staff, who participated in this research, and who so willingly gave us their time, help and support. They also thank Steven Knox and Alan Clelland for their work on programming the mobile phone application. Additional thanks to DynaVox Inc. for supplying the Vmax communication devices to run our system on and Sensory Software Ltd for supplying us with their AAC software. This research was supported by the Research Council UKs Digittal Economy Programme and EPSRC (Grant numbers EP/F067151/1, EP/F066880/1, EP/E011764/1, EP/H022376/1, and EP/H022570 /1).

References

[1] E. Reiter, R. Dale, Building Natural Language Generation Systems, Cambridge University Press, 2000.

[2] J. Light, K. Drager, Aac technologies for young children with complex communication needs: State of the science and future research direc- tions, Augmentative and Alternative Communication 23(3) (2007) 204 – 216.

[3] E. Goldberg, N. Driedger, R. I. Kittredge, Using natural-language processing to produce weather forecasts, IEEE Expert 9(2) (1994) 45 – 53.

[4] E. Reiter, S. Sripada, J. Hunter, I. Davy, Choosing words in computer-generated weather forecasts, Artificial Intelligence 167 (2005) 137 – 169.

[5] A. Gatt, F. Portet, E. Reiter, J. Hunter, S. Mahamood, W. Moncur, S. Sripada, From data to text in the neonatal intensive care unit: Using NLG technology for decision support and information management, AI Communications 22 (2009) 153 – 186.

[6] R. Pérez y Pérez, M. Sharples, Three computer-based models of storytelling: BRUTUS, MINSTREL, and MEXICA, Knowledge-Based Systems 17 (2004) 15 – 29.

[7] R. Black, A. Waller, R. Turner, E. Reiter, Supporting personal narrative for children with complex communication needs, ACM Transactions on Computer-Human Interaction (TOCHI) 19 (2) (2012) 15:1 – 15:35.

[8] J. Reddington, Standing on the shoulders of giants: attacking the meta-problems of technical AAC research, in: Workshop on Speech and Language Processing for Assistive Technologies (SLPAT), Association for Computational Linguistics, 2014.

[9] J. Reddington, The domesday dataset: Linked open data in disability studies, Journal of Intellectual Disabilities 17 (2) (2013) 107 – 121.

[10] A. Waller, Narrative based AAC: From transactional to interactional conversation, in: ISAAC Research Symposium, 2006.

[11] B. Whitmer, OpenBoards - an open communication board format for all AAC users, Last retrieved November 2015, http://www.openboardformat.org.

[12] A. Newell, S. Langer, M. Hickey, The role of natural language processing in alternative and augmentative communication, Natural Language Engineering 4 (1998) 1 – 16.

[13] C. P. McCoy, K., A. Badman, Compansion: From research prototype to practical integration, Natural Language Engineering 4 (1998) 73 – 95.

[14] Y. Netzer, M. Elhadad, Using semantic authoring for blissymbols communication boards, in: HLT, 2006.

[15] R. Manurung, G. Ritchie, H. Pain, A. Waller, D. O' Mara, R. Black, The construction of a pun generator for language skills development., Applied Artificial Intelligence 22(9) (2008) 841 – 869.

[16] A. Waller, R. Black, D. A. O' Mara, H. Pain, G. Ritchie, R. Manurung, Evaluating the standup pun generating software with children with cerebral palsy, ACM Transactions on Accessible Computing 1(3) (2009) 16:27.

[17] J. Todman, A. Alm, D. J. Higginbotham, P. File, Whole utterance approaches in AAC, Augmentative & Alternative Communication 24(3) (2008) 235 – 254.

[18] B. Roark, J. de Villiers, C. Gibbons, M. Fried-Oken, Scanning methods and language modeling for binary switch typing, in: Workshop on Speech and Language Processing for Assistive Technologies, 2010, pp. 28 – 36.

[19] S. Langer, M. Hickey, Using semantic lexicons for full text message retrieval in a communication aid, Natural Language Engineering 4(1) (1998) 41 – 55.

[20] S. S. Nikolova, M. Tremaine, P. R. Cook, Click on bake to get cookies: guiding word-finding with semantic associations, in: ASSETS, 2010, pp. 155 – 162.

[21] G. W. Lesher, D. J. Higginbotham, Using web content to enhance augmentative communication, in: CSUN, 2005.

[22] J. Todman, N. A. Alm, Modelling conversational pragmatics in communication aids, Journal of Pragmatics 3 (2003) 523 – 538.

[23] R. Dye, N. Alm, J. Arnott, G. Harper, A. Morrison, A script-based aac system for transactional interaction, Natural Language Engineering 4(1) (1998) 57 – 71.

[24] R. Schank, R. Abelson, Scripts, plans, goals, and understanding, Lawrence Erlbaum., 1977.

[25] K. F. McCoy, J. Bedrosian, L. Hoag, Implications of pragmatic and cognitive theories on the design of utterance-based aac systems, in: Proceedings of the NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies, 2010, pp. 19 – 27.

[26] J. S. Bruner, From communication to language: A psychological perspective, Cognition 3 (1972) 255 – 289.

[27] J. Kent-Walsh, D. Mcnaughton, Communication partner instruction in aac: Present practices and future directions, Augmentative and Alternative Communication 21 (3) (2005) 195 – 204.

[28] W. Labov, Language in the inner city: Studies in the Black English Vernacular., Philadelphia, University of Pennsylvania Press., 1972.

[29] U. M. Quasthoff K. Nikolaus, What makes a good story? towards the production of conversational narratives, Discourse processing (1982) 16 – 28.

[30] C. Peterson, A. McCabe, Developmental psycholinguistics: Three ways of looking at a childs narrative., New York, Plenum, 1983.

[31] L. Janice, C. Binger, A. K. Smith, Story reading interactions between preschoolers who use aac and their mothers, Augmentative and Alternative Communication 10: 255-268. 10 (1994) 255 – 268.

[32] P. Gervás, Computational approaches to storytelling and creativity, AI Magazine 30(3) (2009) 49 – 62.

[33] C. B. Callaway, J. C. Lester, Narrative prose generation, Artificial Intelligence 139(2) (2002) 213 – 252.

[34] E. Reiter, F. P. A. Gatt, M. van der Meulen., The importance of narrative and other lessons from an evaluation of an NLG system that summarises clinical data., in: International Natural Language Generation Conference (INLG), 2008, pp. 147 – 156.

[35] R. Black, J. Reddington, E. Reiter, N. Tintarev, A. Waller, Using NLG and sensors to support personal narrative for children with complex communication needs, in: Workshop on Speech and Language Processing for Assistive Technologies, 2010, pp. 1 – 9.

[36] R. Black, A. Waller, N. Tintarev, E. Reiter, J. Reddington, A mobile phone based personal narrative system, in: ACM SIGACCESS Conference on Computers and Accessibility, 2011.

[37] L. Chittaro, L. Ieronutti, A visual tool for tracing users' behavior in virtual environments, in: Proceedings of the working conference on Advanced visual interfaces, AVI '04, ACM, New York, NY, USA, 2004, pp. 40 – 47.

[38] K. Sookhanaphibarn, R. Thawonmas, A movement data analysis and synthesis tool for museum visitors' behaviors, in: Advances in Multimedia Information Processing, 2009, pp. 144 – 154.

[39] H. Reichenbach, Elements of Symbolic Logic, Macmillan, 1947.

[40] A. Gatt, E. Reiter, SimpleNLG: A realisation engine for practical applications, in: ENLG, 2009.

[41] J. Reddington, N. Tintarev, Automatically generating stories from sensor data., in: Intelligent User Interfaces, 2011.

[42] R. Flesch, A new readability yardstick, Journal of Applied Psychology 32 (1948) 221 – 233.

[43] TOTeM Labs, Tales of things and electronic memory, Last retrieved November 2015, www.talesofthings.com.

[44] R.Black, R.Turner, E. Reiter, A. Waller., A new tool to support interactive narrative experience for children with communication disorders., in: Conference of the International Society for Augmentative and Alternative Communication, 2010.

[45] J. Reddington, L. Coles-Kemp, Trap hunting: Finding personal data management issues in next generation aac devices, in: Workshop on Speech and Language Processing for Assistive Technologies in association with EMNLP, 2011.

[46] C. Musselwhite, L. Burkhart, Social scripts: co-planned sequenced scripts for AAC users, in: Proceedings of Technology and Persons with Disabilities Conference, 2002.

[47] A. J. Dahlback, N., L. Ahrenberg, Wizard of Oz studies - why and how, Knowledge-Based Systems 6(4) 6(4) (1993) 193 – 200.

[48] D. Boyd, S. Golder, G. Lotan, Tweet, tweet, retweet: Conversational aspects of retweeting on twitter, in: Hawaii International Conference on System Sciences, 2010, pp. 1 – 10. doi:10.1109/HICSS.2010.412.

[49] D. P. Ellis, K. Lee, Minimal-impact audio-based personal archives, in: Workshop on Continuous archival and retrieval of personal experiences in association with the International Multimedia Conference, 2004.

[50] J. Gemmell, L. Williams, K. Wood, R. Lueder, G. Bell, Passive capture and ensuing issues for a personal lifetime store, in: Workshop on Continuous Archival and Retrieval of Personal Experiences in conjunction with the International Multimedia Conference, ACM Press, 2004, pp. 48 – 55.

[51] J. Gemmell, G. Bell, R. Lueder, Mylifebits: a personal database for everything, Commun. ACM 49 (1) (2006) 88 – 95.

[52] S. Mann, Continuous lifelong capture of personal experience with eyetap, in: Workshop on Continuous archival and retrieval of personal experiences in association with the International Multimedia Conference, 2004.

[53] R. Black, J. Reddington, E. Reiter, N. Tintarev, A. Waller, "Hands Busy, Eyes Busy": Generating stories from sensor data for automotive applications, in: Workshop on Multimodal Interfaces for Automotive Applications (MIAA) in conjunction with Intelligent User Interfaces (IUI), 2011.